

## **SOCIOLOGY 703: MULTIVARIATE STATISTICS**

John B. Williamson  
McGuinn 424  
617 552-8530  
Email: JBW@bc.edu

Office Hours:  
Wed, 3:00-3:55  
Wed 5:30-5:55  
Thur, 3:00-4:00  
(or by appointment)

Spring, 2007  
Tues,Thurs @1:30  
McGuinn 030  
TA is Esteban Calvo  
File=Sc703H107.doc

From the major topic headings you will note that we will focus on four statistical procedures: factor analysis, regression analysis, logistic regression, and discriminant analysis. But the course is primarily a course on multiple regression and related procedures. In class I will generally focus on the basic concepts and the most essential material. As this course is designed for students who vary greatly with respect to the strength of their mathematical background and extent to which they expect to be using these skills in the future, I have attempted to build in some flexibility with respect to what you will cover in your assignments. Everyone will be expected to demonstrate competence with respect to the basic material covered in class. This will provide an exposure to what multivariate statistics is all about. Those of you have reasons to go beyond this basic material will be given an incentive (extra credit) for doing so. After you have taken this course you should know when to use each technique, how to set up SPSS runs using the technique, and how to interpret what you see in computer printouts using the technique. I also want you all to be able to critically evaluate journal articles using these techniques. To that end you will be asked to read and critically evaluate published articles using each of these techniques.

### **COURSE REQUIREMENTS**

Approximately every other week I will give an assignment. I strongly urge that you do it by the specified due date. In some cases I will ask you to read and comment on an article that uses a particular statistical procedure. If you turn these assignments in ON TIME, the course TA will provide feedback and suggestions for revision for assignments 1, 2, 4, and 5, but not for assignments 3 or 6 (as they will be due at the same time as the mid-term and the final paper respectively). If after doing an assignment you have doubts about what you did or how you wrote it up, come to see me or the TA, and go over what you did in person. If you have any questions about the comments the TA has made, go see the TA for elaboration or come to see me. If you have doubts about a particular part of an assignment, it might make sense to say so or to ask the TA to take a particularly close look at that part of your write-up. I will suggest that you work on one of the datasets that I have prepared for the course. However, if you come to see me, it will be possible to obtain permission to use your own data at least for some of the assignments. In a folder called Sc703 Data you will find several datasets (GSS2004, PCNORC94, GSS2000, LES101, Wk228); each dataset is an SPSS system file that you will be able to download and use. In the same folder you will also find a word file codebook for each dataset.

## GRADING:

Grading will be based on how much of the material in the lectures, printouts, and reading (including reading covering more advanced issues that we do not have time for in class) that you can convince me that you have learned. When I say how much, I am not suggesting that the more you write the higher the grade. I am saying the more of the most difficult material that you cover in such a way that it is clear to me that you understand it, the higher the grade. I will generally give you a word limit for each assignment. You should allocate the space so that you spend the words showing that you understand the most important aspects of the basic material and as much of the more advanced material as you have time and space for. It is bad strategy to spend most of the space discussing very basic material, if you fully understand some of the more advanced material. As I need to be able to check your work, grading will also be influenced by the clarity not only of the writing, but also the clarity of efforts to present relevant documentation from your SPSS runs (including the relevant material from syntax files) to support assertions in the text of your paper. I need to be able to find that documentation very quickly.

**MID-TERM PAPER** (Due Thursday March 1 @ 3:00 pm). This paper will consist of Assignments 1, 2, and 3. It will make up 40% of your grade for the course.

**FINAL PAPER** (Due Thursday May 10th @ 3:00 pm). This paper will consist of Assignments 4, 5, and 6. The Final Paper will make up 60% of your grade.

## COURSE TEXTS:

Polit, Denise F. 1996. *Data analysis & Statistics for Nursing Research*. Stanford, CT: Appleton & Lange. This text is required. (I have ordered copies for the bookstore and it will also be in the library. This will be your source for a basic treatment of the topic)

Tabachnick, Barbara G., & Linda S. Fidell. 2007. *Using Multivariate Statistics*. Fifth Edition. Boston: Pearson/Allyn and Bacon. This text is suggested and not required, but I have ordered it at the bookstore. This is the place to go for a treatment that is more thorough and more sophisticated than what you will find in Polit or in Noursis. If you suspect or know that you will be doing quantitative research after you leave this course, this book (which is expensive) is probably worth the cost. One unusual aspect of the book is that an effort is made to introduce the reader to matrix algebra and the use of matrix algebra to explain what is going on. As you will see, this does not mean that you must know or learn matrix algebra to use this book (because they only use matrix algebra in advanced sections that you can skip.) I have compared this 5<sup>th</sup> edition with the prior (2001) 4<sup>th</sup> edition and the differences are quite minor. If you have access to a 4<sup>th</sup> edition do not hesitate to use it, but keep in mind that in the prior edition logistic regression is covered in Chapter 10 (not 12) and discriminant analysis is covered in Chapter 9 (not 11).

Norusis, Marija J. 2005. *SPSS 13.0 Statistical Procedures Companion*. Upper Saddle River, NJ: Pearson/Prentice Hall. This will be a very useful book as it will discuss various statistics that are presented in your SPSS output. If you buy on line, you may want to get a version that is more recent than this. I have selected this as it

correspondence to the current version of SPSS, but you never know when they will update BC to a more current version. It does not matter much as the changes from version to version are generally quite modest.

### TOPIC AND READING:

Items on reserve at O'Neill are indicated by (O'Neill) at the end of the citation. Items marked as (B.C. Stacks) are not on reserve and are listed primarily for future reference.

Most if not all of the articles that are assigned can be downloaded from the internet if you like. Check with the reserve room or the course TA to find out how to download articles.

**Items that everyone will read are marked with an asterisk.** Each of you will read items without an asterisk selectively depending on whether you are looking for a more basic or a more advanced treatment and depending on which version of SPSS you are working with. **You will always want to read at least one of the SPSS sources in connection with each procedure we cover. I recommend that you start with *SPSS 13.0 Statistical Procedures Companion*.**

I will be leaving it up to you to decide which of the various sources to use in connection with each procedure and topic. It will depend on how sophisticated a treatment you want. Most of you will eventually figure that out and do most of your basic reading from just two or possibly three sources. When I cover a technique, concept, statistic, option, etc. after class look it up and read about it in at least one of the sources that I have provided. If you are having trouble finding a relevant discussion see me or the TA for the course.

### FACTOR ANALYSIS

**OBJECTIVES:** To be able to do factor analysis using SPSS. By the time we finish this topic you should understand such terms as: common factors, communality, eigenvalue, scree plot, factor loadings, factor pattern matrix, factor structure matrix, factor rotation, orthogonal rotation, oblique rotation, factor scores, factor score coefficient matrix, principal components analysis, principal axis factoring, plots of factor loadings, reproduced correlation matrix, and anti-image correlation matrix. You should be able to interpret an SPSS syntax file for a factor analysis run and you should be able to write a SPSS factor analysis program using a syntax file. When doing this and all other assignments, please turn in the appropriate syntax file corresponding to any and all SPSS output presented. I think that you will find that it is generally best to extract the tables as object files from the SPSS output file using “copy objects” and paste them into a word document. In your reports before each set of tables paste in the syntax used to generate those tables (or present a complete syntax file for the assignment in an appendix).

**READING:**

\*Polit, Denise F. 1996. *Data Analysis & Statistics for Nursing Research*, Chapter 12. Stanford, CT: Appleton & Lange. (This is a simple introduction to factor analysis.) (O'Neill)

\*Norusis, Marija J. 2005. Chapter 17 (Factor Analysis) in *SPSS 13.0 Statistical Procedures Companion*. This is a good statistical overview of factor analysis and what you get when you run it in SPSS, but it does not say anything about how to set up your run. If they do not have version 13.0, ask for version 12.0 (2003). (O'Neill), (GSSW).

\*Mertler, Craig A., & Rachel A. Vannatta. 2005. *Advanced and Multivariate Statistical Methods*. Third Edition. See Ch. 9. This book provides an alternative to the Polit text written at about the same level. You may prefer it. You should try it out. This is a good place to compare it with Polit. (O'Neill)

Norusis, Marija J. 1990. *SPSS Advanced Statistics Student Guide*, Chapter 5, Chicago: SPSS. (This treatment is based on a dated version of SPSS, but it has a more thorough explanation of the statistical aspects of factor analysis than I find in many of the more recent SPSS manuals. This treatment is at about the level of the Polit book.) (O'Neill)

SPSS. 1999. *SPSS Base 10.0 Applications Guide*, Chapter 16. Chicago: SPSS. (This is a good overview of factor analysis as it is presented in version 10.0 of SPSS. An effort is made to explain a number of the key statistics presented.) (O'Neill).

Tabachnick, Barbara G., & Linda S. Fidell. 2007. *Using Multivariate Statistics*, Fifth Edition. Chapter 13. Boston: Pearson/Allyn & Bacon. This is a more advanced treatment of factor analysis than you will find in Polit. You will need to know (or learn) some matrix algebra to fully comprehend some parts of this chapter. For a brief introduction see Appendix A of the book. For a more thorough introduction see *Matrix Algebra: An Introduction* by Krishnan Namboodiri (Sage, 1984, on reserve at O'Neill). While you need to understand matrix algebra to understand some sections of this chapter, the other sections can be read without understanding matrix algebra. There is not loss of continuity if you just skip some of the more advanced material based on the matrix algebra. If they do not have the 5<sup>th</sup> edition ask for the 4<sup>th</sup> edition (2001). (O'Neill)

Spicer, John. 2005. *Making Sense of Multivariate Data Analysis* (Thousand Oaks, CA: Sage). This is written at a basic level and makes minimal use of numbers. It provides a good overview and even discusses confirmatory factor analysis, but all at a relatively elementary level. See Ch. 7 (O'Neill).

\*Hibbs, Douglas A. 1973. Ch. 2, "Mass Political Violence--Its Scope and Measurement." (This is an example of the use of factor analysis in published research.) (This will be on the course CD).

\*Martin, Greg et al. 2006. "The Adolescent Cannabis Problem Questionnaire (CPQ-A): Psychometric Properties." *Addictive Behaviors* 31: 2238-2248. .)(This will be on the course CD)

Laroche, Michael, et al. 1997. "A Comparative Investigation of Dimensional Structures of Acculturation for Italian Canadians and Greek Canadians." *Journal of Social Psychology* 137 (3): 317-331. This article makes use of such concepts as confirmatory and exploratory factor analysis. It uses LISREL models. This article is a bit advanced for this course, but it will give you some idea of what a more advanced version of factor analysis can look like. (B.C. Stacks)

## **MULTIPLE REGRESSION AND RELATED TECHNIQUES**

**OBJECTIVES:** To be able to do multiple regression and path analysis using SPSS. After we finish this section of the course you should understand such concepts as: slope, intercept, least squares line, simple vs. multiple regression, regression coefficient vs. partial regression coefficient, the multiple correlation coefficient, the coefficient of determination, R-square, adjusted R-square, unstandardized partial regression coefficients, standardized partial regression coefficients, beta weights, t values, hierarchical multiple regression, stepwise multiple regression, forward inclusion, backward elimination, part correlation, partial correlation, collinearity and multicollinearity, tolerance, residual analysis, standardized residuals, assumptions about residuals (normality, linearity, homoscedasticity), partial plots, the construction and use of dummy variables, interaction terms and the analysis of interaction, data transformation, quadratic regression, path analysis, path coefficients, direct effects, indirect effects, spurious effects, causal analysis, path diagrams.

## **READING**

\*Polit, Denise F. 1996. *Data Analysis & Statistics for Nursing Research*. Stanford, CT: Appleton & Lange. Chapter 9 Correlation and Simple Regression (much of this will be a review of material covered in Sc702), Chapter 10 Multiple Regression (a good basic introduction to multiple regression), Chapter 14 Causal Modeling: Path Analysis and Linear Structural Relations Analysis (a basic introduction to some advanced topics; treatment toward the end of the chapter may be too brief to be of much use). (O'Neill)

\*Norusis, Marija J. 2005. *SPSS 13.0 Statistical Procedures Companion*. Chapter 12 (Bivariate Linear Regression) and Chapter 13 (Multiple Linear Regression). This

is a good statistical overview of regression analysis and what you get when you run it in SPSS, but it does not say anything about how to set up your run. If they do not have version 13.0, ask for version 12.0 (2003). (O'Neill), (GSSW).

Mertler, Craig A., & Rachel A. Vannatta. 2005. *Advanced and Multivariate Statistical Methods*. Third Edition. Glendale, CA: Pycszak Publishing. This book is suggested, but not required. I have ordered copies for the bookstore. If you want more detail than you find in the Polit book, this is a good place to go. It will be easier than Tabachnick and Fidell. I recommend Ch 3 on Data Screening, Ch. 7 on Multiple Regression, and Ch 8 on Path Analysis. (O'Neill)

SPSS. 1999. *SPSS Base 10.0 Applications Guide*, Chapter 12. Chicago: SPSS. (This is a good overview of how to do simple and multivariate regression in SPSS. There is also some discussion of residual analysis.) (O'Neill)

Norusis, Marija J. 2002. *SPSS 11.0 Guide to Data Analysis*, Chapters 20-24. Upper Saddle River, NJ: Prentice Hall. (This is a good overview of the material on regression that we will be covering for SPSS with an effort to explain the statistics involved.) (O'Neill) (GSSW)

Bohrstedt, George W., & David Knoke. 1994. *Statistics for Social Data Analysis*, 3<sup>rd</sup> Edition. Itasca, IL: F.E. Peacock Publishers. (This is a basic discussion of nonlinear relationships and data transformations. You will find that by mistake the order in which the two authors are listed has been reversed on the cover of the book). (O'Neill)

McClendon, McKee J. 1994. *Multiple Regression and Causal Analysis*. Itasca, IL: F.E. Peacock Publishers. Chapter 3 Multiple Regression (Good basic treatment of topic), Chapter 5 Nominal Independent Variables (an advanced version of dummy variable analysis), Chapter 7 Nonadditive Relationships (good treatment of interaction), Chapter 8 Causal Analysis I (a good discussion of path analysis). (O'Neill)

Tabachnick, Barbara G., & Linda S. Fidell. 2007. *Using Multivariate Statistics*, Fifth Edition. Chapter 13. Boston: Pearson/Allyn & Bacon. Chapter 4 Cleaning Up Your Act: Screening Data Prior to Analysis (good discussion of various issues related to getting the data ready for statistical analysis), Chapter 5 Multiple Regression (a place to go for a more advanced and more detailed treatment of the topic than you will find in the Polit book), Chapter 14 Structural Equation Modeling (optional advanced material, a place to go for a bit more advanced treatment of multiple regression than we have time for in Sc703). . If they do not have the 5<sup>th</sup> edition ask for the 4<sup>th</sup> edition (2001). This book a more advanced treatment of multiple regression than you will find in Polit. You will need to know (or learn) some matrix algebra to fully comprehend some parts of these chapters. For a brief introduction see Appendix A of the book. For a more

thorough introduction see *Matrix Algebra: An Introduction* by Krishnan Namboodiri (Sage, 1984 also on reserve at O'Neill). (O'Neill)

Spicer, John. 2005. Ch. 4 of *Making Sense of Multivariate Data Analysis* (Thousand Oaks, CA: Sage). This is written at a basic level and makes minimal use of numbers. (O'Neill)

Keith, Timothy K. 2006. *Multiple Regression and Beyond*. Pearson Education: Boston. This book provides a thorough treatment of many of the issues we cover in this course. It is not written at an overly sophisticated level. I urge you to spend some time with it, you might find this book very useful for this course and for future work using multiple regression, path analysis, and related procedures. (O'Neill)

\*Shen, Ce. & John B. Williamson. 1999. "Maternal Mortality, Women's Status, and Economic Dependency in Less Developed Countries: A Cross-National Analysis. *Social Science & Medicine* 49:197-214. (This is an article illustrating the use of multiple regression analysis and path analysis.) (O'Neill)

\*Pampel, Fred C., John B. Williamson, and Robin Stryker. 1990. "Class Context and Pension Response to Demographic Structure in Advanced Industrial Democracies." *Social Problems* 37:535-550. (This is an example of multiple regression analysis checking for and reporting interaction effects.) (O'Neill)

## LOGISTIC REGRESSION

**OBJECTIVES:** To be able to do logistic regression using SPSS. After we finish this section of the course you should understand such concepts as: the logistic regression curve, logit coefficients, standard error of the logit coefficient, Wald statistic, partial correlation (the special version used in logistic regression), odds ratio, and assessing goodness of fit of the model.

## READING

\*Polit, Denise F. 1996. *Data Analysis & Statistics for Nursing Research*, Chapter 13, pp. 389-412. Stamford, CT: Appleton & Lange. (This is a simple introduction to Logistic Regression.) (O'Neill)

\*Norusis, Marija J. 2005. Chapter 15 (Logistic Regression Analysis) in *SPSS 13.0 Statistical Procedures Companion*. This is a good statistical overview of logistic regression and what you get when you run it in SPSS, but it does not say anything about syntax describing how to set up your run. . If they do not have version 13.0, ask for version 12.0 (2003). (O'Neill), (GSSW).

\*Mertler, Craig A., & Rachel A. Vannatta. 2005. *Advanced and Multivariate Statistical Methods*. Third Edition. See Ch. 11. (O'Neill)

SPSS. 2001. *SPSS Regression Models 11.0*. Chicago: SPSS. See Chapter 2 (Logistic Regression) and Chapter 8 (Logistic Regression Analysis Examples). A good place to go after you read Polit. (O'Neill) (GSSW)

Tabachnick, Barbara G., & Linda S. Fidell. 2007. *Using Multivariate Statistics*, Fifth Edition. Chapter 10. Boston: Pearson/Allyn & Bacon. (O'Neill). If they do not have the 5<sup>th</sup> edition, try the 4<sup>th</sup> edition Chapter 10 Logistic Regression. This is a more advanced treatment of logistic regression than you will find in Polit. You will need to know (or learn) some matrix algebra to fully comprehend some parts of this chapter. For a brief introduction see Appendix A of the book. For a more thorough introduction see *Matrix Algebra: An Introduction* by Krishnan Namboodiri (Sage, 1984, also on reserve at O'Neill). (O'Neill)

Bohrstedt, George W., & David Knoke. 1994. *Statistics for Social Data Analysis*, 3<sup>rd</sup> Edition. Itasca, IL: F.E. Peacock Publishers. Chapter 9, pp. 334-359. (This is a good discussion of logistic regression that is a bit more advanced than Polit.) (O'Neill)

Spicer, John. 2005. *Making Sense of Multivariate Data Analysis* (Thousand Oaks, CA: Sage). This is written at a basic level and make minimal use of numbers. It provides a good overview of logistic regression, but at a relatively elementary level. See Ch. 5. (O'Neill).

SPSS 11.0 *Syntax Reference Guide*. 2001 Chicago: SPSS, Inc. This gives the details of SPSS syntax for this procedure and all of the other procedures that we consider in this course. (O'Neill)

Gay, David, and John Lynxwiler. 1999. "The Impact of Religiosity on Race Variation in Abortion Attitudes." *Sociological Spectrum* 19:359-377. (This is a simple example of logistic regression.) (O'Neill)

Jones, Rachel K., and Ye Luo. 1999. "The Culture of Poverty and African-American Culture: An Empirical Assessment." *Sociological Perspectives* 42: 439-458. (This is a simple example of logistic regression.) (O'Neill)

## DISCRIMINANT ANALYSIS

**OBJECTIVES:** To be able to do discriminant analysis (also called discriminant function analysis) using SPSS. You should understand such terms as: discriminant function, discriminant score, group centroids, eigenvalue, Wilks' lambda, canonical correlation, unstandardized canonical discriminant function coefficients, standardized canonical discriminant function coefficients, and structure matrix

## READING

\*Polit, Denise F. 1996. *Data Analysis & Statistics for Nursing Research*, Chapter 14. Stanford, CT: Appleton & Lange. (This is a simple introduction to discriminant analysis.) (O'Neill)

\*Norusis, Marija J. 2005. Chapter 14 (Discriminant Analysis) in *SPSS 13.0 Statistical Procedures Companion*. This is a good statistical overview of discriminant analysis and what you get when you run it in SPSS, but it does not say anything about how to set up your run. . If they do not have version 13.0, ask for version 12.0 (2003). (O'Neill), (GSSW).

\*Mertler, Craig A., & Rachel A. Vannatta. 2005. *Advanced and Multivariate Statistical Methods*. Third Edition. See Ch. 10. (O'Neill)

SPSS. 1999. *SPSS Base 10.0 Applications Guide*, Chapter 14 Discriminant Analysis. Chicago: SPSS. (This is a good overview of how to do discriminant analysis in SPSS. (O'Neill)

Spicer, John. 2005. *Making Sense of Multivariate Data Analysis* (Thousand Oaks, CA: Sage). This is written at a basic level and make minimal use of numbers. It provides a good overview of discriminant analysis, but at a relatively elementary level. See Ch. 5. (O'Neill).

SPSS 11.0 *Syntax Reference Guide*. 2001 Chicago: SPSS, Inc. This gives the details of SPSS syntax for this procedure and all of the other procedures that we consider in this course. (O'Neill)

SPSS. 2002. *SPSS Base 11.5 User's Guide*, Chapter 28 Discriminant Analysis. Chicago: SPSS. (This is a very brief treatment of discriminant analysis with a focus on how to run the program and what you get.) (O'Neill) (GSSW)

Tabachnick, Barbara G., & Linda S. Fidell. 2007. *Using Multivariate Statistics*, Fifth Edition. Chapter 9. Boston: Pearson/Allyn & Bacon. (O'Neill). If they do not have the 5<sup>th</sup> edition, try the 4<sup>th</sup> edition Chapter 11. This is a more advanced treatment of logistic regression than you will find in Polit. You will need to know (or learn) some matrix algebra to fully comprehend some parts of this chapter. For a brief introduction see Appendix A of the book. For a more thorough introduction see *Matrix Algebra: An Introduction* by Krishnan Namboodiri (Sage, 1984, also on reserve at O'Neill). (O'Neill)

Dattalo, Patrick. 1994. "A Comparison of Discriminant Analysis and Logistic Regression." *Journal of Social Service Research* 19 (3/4):121-144. (An example of an article using discriminant analysis). (O'Neill)

Sightler, Kevin W. & Janet S. Adams. 1999. "Differences between Stayers and Leavers among Part-Time Workers." *Journal of Managerial Issues* 11(1): 110-125. (An example of discriminant analysis) (O'Neill).

### **Suggestions for Mid-Term and Final Papers (as well as assignments 1-6)**

1. You are free to submit your original assignments without making changes or to make some revisions. It is not a problem if there are some comments from the TA on the version that you submit to be graded by me.
2. It is very important that you include the SPSS tables (that the TA and I will need to check what you have done). Do not include tables that you do not mention in your report. It works best to use cut and paste function (copy objects) to insert tables into the body of your write-up the segment at the appropriate place. It is important to make it very easy for us to find the relevant material.
3. I will give you a maximum word limit for each assignment. If I specify a word limit and only half of the class follows it, we have a problem. It is unfair to those who would have found it easy to say more and would like to have had the space to do so. If I do not set a word limit some people will write very long papers and when others notice that, everyone's papers start to get very long (increasing the time everyone has to spend on these assignments). **How do I deal with these two possible outcomes?** My strategy this year is going to be to require that you stay under the specified word limit and that you report both the upper word limit and the actual length of your assignment on the first page of your assignment. Use the word count tool in MS Word to do the actual counting.
4. **The most important time during the course to avoid getting behind is toward the end, specifically when doing the last couple of assignments. They will be the most difficult, will probably take longer than others, and will count the most. By that time in the semester the course will seem to be moving very quickly and most students will experience increased time demands from other courses. The solution: be sure to keep up, particularly toward the end of the course.**
5. **Does it make sense to try to show that you have learned some tricks that we did not cover in class?** If you have the time to do some of the optional supplementary reading and if you understand what you have read and if you can find a way to do some computer runs using this new material, then be sure to find a way to show this in your assignments. If you elect to learn optional advanced material, it is up to you to learn this new material on your own (do not expect the TA to teach it to you). If you do so and understand what you have read, then there will be some extra credit. I suggest that you find a way to work in some additional runs using some tricks (options, etc.) that we did not cover in class, options that are not covered in the runs we have (or will have) gone over in class. Be sure to include a discussion of the runs to illustrate that you understand how to interpret what you have found. There is no extra credit for just doing the run that does something new. Also there is no extra credit for doing something early, that is, doing it before we get to it in class. If you are not sure whether or not we are going to cover the topic in class later in the course, please ask.